



Orchestrate Workflows with the Data Agent Kit

Course description

Orchestrate Workflows with the Data Agent Kit is an advanced-level, instructor-led course designed for practitioners who want to bring natural-language and AI-driven workflows into modern data engineering environments. The course introduces the foundations of agentic data engineering and the Data Agent Kit architecture. Participants learn how ADK-based data agents can interact with Google Cloud data services and how the Model Context Protocol (MCP) provides connectivity between agents and data platforms.

Learners build and configure a data agent environment, connect to BigQuery, Spanner, and Cloud SQL, and use the agent to discover and understand enterprise data. The course then explores natural-language-driven data engineering tasks including catalog browsing, universal search, schema and lineage inspection, data quality analysis, SQL transformation generation, BigQuery DataFrames, notebooks, and medallion architecture.

Participants also learn how to use the data agent to orchestrate pipelines, build DAGs using natural language, schedule deployments, monitor pipelines, and diagnose failures. The course extends into machine learning workflows using BigQuery ML, including model training, inference, and drift monitoring. Finally, learners explore enterprise security practices including least privilege, principal access boundaries, IAM deny policies, prompt injection protection, document guardrails, and Model Armor.

- Course level: Advanced
- Duration: 5 hours 45 minutes

Activities

This course includes instructor-led presentations, technical demonstrations, guided data agent development, hands-on ADK exercises, MCP integration activities, natural-language data engineering and data discovery scenarios, data transformation exercises, pipeline orchestration and troubleshooting workflows, BigQuery ML demonstrations, security and governance scenarios, architecture discussions, knowledge checks, and a final quiz.

Course objectives

In this course, you will learn to:

- Explain the foundations of agentic data engineering and describe the architecture of the Data Agent Kit.
- Explain how the Data Agent Kit integrates with the Agent Development Kit (ADK).
- Build a basic ADK-based data agent and configure the Data Agent Kit development environment.
- Configure authentication and MCP connectivity, and apply service account impersonation for secure access.
- Connect data agents to BigQuery, Spanner, and Cloud SQL.
- Discover and explore enterprise data using natural-language interactions.
- Browse data catalogs and perform universal search across available data resources.
- Inspect schemas, lineage, and data quality using natural-language interactions.
- Generate data transformations for business data pipelines and organize data using the medallion architecture.
- Generate SQL transformations using natural-language prompts, and work with BigQuery DataFrames and notebooks.
- Orchestrate data pipelines using the data agent and build DAGs using natural language.
- Schedule and deploy pipelines using GitHub Actions, and monitor pipeline execution through the agent.
- Resolve data pipeline incidents and trace data lineage to identify reconciliation gaps.
- Diagnose pipeline failures using agent-driven workflows.

- Build and train models using BigQuery ML, perform inference, and monitor model drift.
- Secure agent environments using enterprise security controls, including least-privilege access with the MCP Tool User role.
- Configure principal access boundaries and apply IAM deny policies for MCP.
- Protect agents against prompt injection, apply document guardrails, and configure Model Armor to support secure agent workflows.

Intended audience

This course is intended for professionals building AI-driven data engineering workflows on Google Cloud:

- Data Engineers
- Data Scientists
- Software Developers and Engineers
- AI/ML Engineers
- Cloud Engineers
- Platform Engineers
- DevOps Engineers
- MLOps Engineers
- Technical Architects
- Solutions Architects
- Data Architects
- Professionals building AI-driven data engineering workflows

Prerequisites

We recommend that attendees of this course have:

- Familiarity with Google Cloud data services is recommended.
- Familiarity with data engineering concepts is beneficial.
- Familiarity with SQL and data transformation concepts is recommended.
- Basic understanding of the Agent Development Kit (ADK) is beneficial.
- Familiarity with MCP concepts is helpful.
- Experience with IDE and CLI-based development workflows is recommended.
- Basic understanding of BigQuery is beneficial.
- Familiarity with cloud security and IAM concepts is recommended.

Course outline

Module 1: Foundations of Agentic Data Engineering

- The role of agentic AI in modern data engineering
- Scenario setup with Cymbal Superstore and the Cymbal discount initiative
- Data engineering agent interfaces
- The Data Agent Kit architecture and supporting ADK patterns
- Demo: Data engineering agent interfaces and ADK patterns.
- Scenario walkthrough: Cymbal Superstore, with architecture walkthrough and knowledge check.

Module 2: Build Your Data Agent Environment

- Building and installing the development environment
- Building a basic ADK data agent
- Installing the Data Agent Kit starter pack
- Configuring the Data Agent Kit VS Code extension
- Hands-on Lab: Build and install the environment and create a basic ADK data agent.
- Demo: Data Agent Kit starter pack and VS Code extension configuration.

Module 3: Authentication and MCP Connectivity

- Authentication for data agent workflows and establishing MCP connectivity
- Service account impersonation
- Local and remote MCP servers
- How MCP enables agents to interact with data services
- Hands-on Lab: Service account impersonation.
- Guided activity: Connect data agents to BigQuery, Spanner, and Cloud SQL.

Module 4: Discover and Explore Data

- Discovering Cymbal Superstore data resources
- Browsing the data catalog with the data agent
- Universal search across available data resources
- Chat-driven schema, lineage, and data quality inspection
- Hands-on Lab: Schema and lineage inspection.
- Demo: Universal search, catalog browsing, and data quality analysis.

Module 5: Generate Data Transformations

- Generating transformations for the discount pipeline
- Organizing data using the medallion architecture

- Generating SQL transformations using natural-language interactions
- BigQuery DataFrames and notebooks
- Hands-on Lab: Generate SQL transformations.
- Demo: Medallion architecture, BigQuery DataFrames, and notebooks.

Module 6: Orchestrate Pipelines

- Pipeline orchestration with the Data Agent Kit
- Building DAGs with natural-language instructions
- Scheduling pipeline execution and GitHub Actions deployment
- Monitoring pipeline execution through the agent
- Hands-on Lab: Build a DAG using natural language.
- Demo: Scheduling, GitHub Actions deployment, and monitoring walkthrough.

Module 7: Resolve Pipeline Incidents

- Investigating pipeline incidents with the data agent
- Tracing lineage to resolve reconciliation gaps
- Diagnosing pipeline failures
- Relationships between pipeline execution, lineage, and data issues
- Hands-on Lab: Investigate reconciliation gaps.
- Scenario: Resolve a pipeline incident and diagnose pipeline failures.

Module 8: Model Discount Elasticity with BigQuery ML

- Modeling discount elasticity with BigQuery ML
- Training the discount elasticity model within the data engineering workflow
- Inference using the trained model
- Model drift monitoring
- Guided activity: Train the discount elasticity model.
- Demo: Model inference and drift monitoring.

Module 9: Secure Agent Environments

- Securing environments for data agents
- Applying least privilege with the MCP Tool User role
- Principal access boundaries and IAM deny policies for MCP
- The prompt injection threat model and guarding against prompt injection
- Document guardrails and securing data used by agent workflows
- Configuring Model Armor
- Hands-on Lab: Least-privilege configuration.

- Demo: Principal access boundaries, IAM deny policies, document guardrails, and Model Armor.

Module 10: Course Quiz and Completion

- Course review and knowledge validation
- Reinforcing agentic data engineering workflows, MCP connectivity, and data service integration
- Reviewing pipeline orchestration, troubleshooting, security, and governance practices
- Scenario-based knowledge check and final quiz.

Your course overview ends here.
But your cloud journey doesn't!

Unlock advanced training and certification pathways with **CloudThat**.



About CloudThat

Established in 2012, CloudThat is an award-winning company and the first in India to offer cloud training and consulting services for individuals and enterprises worldwide. Recently, it won Google Cloud's New Training Partner of the Year Award for 2025, becoming the first company in the world in 2025 to hold awards from all three major cloud giants: AWS, Microsoft, and Google. CloudThat notably won consecutive AWS Training Partner of the Year (APJ) awards in 2023 and 2024 and the Microsoft Training Services Partner of the Year Award in 2024, bringing its total award count to an impressive 12 awards in the last 8 years.

CloudThat specializes in Cloud Migration, Data Platforms, DevOps, IoT, and cutting-edge technologies like Gen AI & AI/ML. As a Microsoft Solutions Partner, AWS Premier Tier Training Partner, Google Cloud Platform Partner, and collaborator with leading organizations like HPE, Oracle, Nvidia and Databricks, CloudThat has trained over 850,000 professionals across 600+ cloud certifications, empowering students and professionals worldwide to advance their skills and careers.

To know more about our cloud certification training, email at sales@cloudthat.com or call us at [+91-8880002200](tel:+91-8880002200) or [+1-8555588830](tel:+1-8555588830) for those from outside India.